

1600

1600

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20:14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

ENTERED

3 <110> APPLICANT: Gijzen, Mark  
5 <120> TITLE OF INVENTION: Soybean Seed Coat Peroxidase Structural Gene And Regulatory

Region

7 <130> FILE REFERENCE: 76-105  
9 <140> CURRENT APPLICATION NUMBER: US 08/939,905D  
C--> 11 <141> CURRENT FILING DATE: 1997-09-29  
13 <150> PRIOR APPLICATION NUMBER: US 08/723,414  
15 <151> PRIOR FILING DATE: 1996-09-30  
17 <160> NUMBER OF SEQ ID NOS: 20  
19 <170> SOFTWARE: PatentIn version 3.0  
21 <210> SEQ ID NO: 1  
23 <211> LENGTH: 1244  
25 <212> TYPE: DNA  
27 <213> ORGANISM: Glycine max  
29 <220> FEATURE:  
31 <221> NAME/KEY: CDS  
33 <222> LOCATION: (1)..(1056)  
35 <220> FEATURE:  
37 <221> NAME/KEY: sig\_peptide  
39 <222> LOCATION: (1)..(78)  
42 <400> SEQUENCE: 1 48  
43 atg ggt tcc atg cgt cta tta gta gtg gca ttg ttg tgt gca ttt gct 48  
44 Met Gly Ser Met Arg Leu Leu Val Val Ala Leu Leu Cys Ala Phe Ala  
45 1 5 10 15 96  
47 atg cat gca ggt ttt tca gtc tct tat gct cag ctt act cct acg ttc 96  
48 Met His Ala Gly Phe Ser Val Ser Tyr Ala Gln Leu Thr Pro Thr Phe  
49 20 25 30 144  
51 tac aga gaa aca tgt cca aat ctg ttc cct att gtg ttt gga gta atc 144  
52 Tyr Arg Glu Thr Cys Pro Asn Leu Phe Pro Ile Val Phe Gly Val Ile  
53 35 40 45 192  
55 ttc gat gct tct ttc acc gat ccc cga atc ggg gcc agt ctc atg agg 192  
56 Phe Asp Ala Ser Phe Thr Asp Pro Arg Ile Gly Ala Ser Leu Met Arg  
57 50 55 60 240  
59 ctt cat ttt cat gat tgc ttt gtt caa ggt tgt gat gga tca gtt ttg 240  
60 Leu His Phe His Asp Cys Phe Val Gln Gly Cys Asp Gly Ser Val Leu  
61 65 70 75 80 288  
63 ctg aac aac act gat aca ata gaa agc gag caa gat gca ctt cca aat 288  
64 Leu Asn Asn Thr Asp Thr Ile Glu Ser Glu Gln Asp Ala Leu Pro Asn  
65 85 90 95 336  
67 atc aac tca ata aga gga ttg gac gtt gtc aat gac atc aag aca gcg 336  
68 Ile Asn Ser Ile Arg Gly Leu Asp Val Val Asn Asp Ile Lys Thr Ala  
69 100 105 110 384  
71 gtg gaa aat agt tgt cca gac aca gtt tct tgt gct gat att ctt gct 384  
72 Val Glu Asn Ser Cys Pro Asp Thr Val Ser Cys Ala Asp Ile Leu Ala  
73 115 120 125 432  
75 att gca gct gaa ata gct tct gtt ctg gga gga ggt cca gga tgg cca 432  
76 Ile Ala Ala Glu Ile Ala Ser Val Leu Gly Gly Pro Gly Trp Pro  
77 130 135 140

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20:14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

79	gtt	cca	tta	gga	aga	agg	gac	agc	tta	aca	gca	aac	cga	acc	ctt	gca	480		
80	Val	Pro	Leu	Gly	Arg	Arg	Asp	Ser	Leu	Thr	Ala	Asn	Arg	Thr	Leu	Ala			
81	145									155					160				
83	aat	caa	aac	ctt	cca	gca	cct	ttc	ttc	aac	ctc	act	caa	ctt	aaa	gct	528		
84	Asn	Gln	Asn	Leu	Pro	Ala	Pro	Phe	Phe	Asn	Leu	Thr	Gln	Leu	Lys	Ala			
85										170					175				
87	tcc	ttt	gct	gtt	caa	ggt	ctc	aac	acc	ctt	gat	tta	gtt	aca	ctc	tca	576		
88	Ser	Phe	Ala	Val	Gln	Gly	Leu	Asn	Thr	Leu	Asp	Leu	Val	Thr	Leu	Ser			
89										185					190				
91	ggt	ggt	cat	acg	ttt	gga	aga	gct	cgg	tgc	agt	aca	ttc	ata	aac	cga	624		
92	Gly	Gly	His	Thr	Phe	Gly	Arg	Ala	Arg	Cys	Ser	Thr	Phe	Ile	Asn	Arg			
93										200					205				
95	tta	tac	aac	ttc	agc	aac	act	gga	aac	cct	gat	cca	act	ctg	aac	aca	672		
96	Leu	Tyr	Asn	Phe	Ser	Asn	Thr	Gly	Asn	Pro	Asp	Pro	Thr	Leu	Asn	Thr			
97										215					220				
99	aca	tac	tta	gaa	gta	ttg	cgt	gca	aga	tgc	ccc	cag	aat	gca	act	ggg	720		
100	Thr	Tyr	Leu	Glu	Val	Leu	Arg	Ala	Arg	Cys	Pro	Gln	Asn	Ala	Thr	Gly			
101										230					235				
103	gat	aac	ctc	acc	aat	ttg	gac	ctg	agc	aca	cct	gat	caa	ttt	gac	aac	768		
104	Asp	Asn	Leu	Thr	Asn	Leu	Asp	Leu	Ser	Thr	Pro	Asp	Gln	Phe	Asp	Asn			
105										245					250				
107	aga	tac	tac	tcc	aat	ctt	ctg	cag	ctc	aat	ggc	tta	ctt	cag	agt	gac	816		
108	Arg	Tyr	Tyr	Ser	Asn	Leu	Leu	Gln	Leu	Asn	Gly	Leu	Leu	Gln	Ser	Asp			
109										260					265				
111	caa	gaa	ctt	ttc	tcc	act	cct	ggt	gct	gat	acc	att	ccc	att	gtc	aat	864		
112	Gln	Glu	Leu	Phe	Ser	Thr	Pro	Gly	Ala	Asp	Thr	Ile	Pro	Ile	Val	Asn			
113										275					280				
115	agc	ttc	agc	agt	aac	cag	aat	act	ttc	ttt	tcc	aac	ttt	aga	gtt	tca	912		
116	Ser	Phe	Ser	Ser	Asn	Gln	Asn	Thr	Phe	Phe	Ser	Asn	Phe	Arg	Val	Ser			
117										290					295				
119	atg	ata	aaa	atg	gtt	aat	att	gga	gtg	ctg	act	ggg	gat	gaa	gga	gaa	960		
120	Met	Ile	Lys	Met	Gly	Asn	Ile	Gly	Val	Leu	Thr	Gly	Asp	Glu	Gly	Glu			
121										305					310				
123	att	cgc	ttg	caa	tgt	aat	ttt	gtg	aat	gga	gac	tcg	ttt	gga	tta	gtc	1008		
124	Ile	Arg	Leu	Gln	Cys	Asn	Phe	Val	Asn	Gly	Asp	Ser	Phe	Gly	Leu	Ala			
125										325					330				
127	agt	gtg	gct	tcc	aaa	gat	gct	aaa	caa	aag	ctt	gtt	gct	caa	tct	aaa	1056		
128	Ser	Val	Ala	Ser	Lys	Asp	Ala	Lys	Gln	Lys	Leu	Val	Ala	Gln	Ser	Lys			
129										340					345				
131	taa	acc	aata	at	aat	gggg	at	gt	gc	at	gt	ca	taa	agg	caa	at	tag	ttgt	1116
133	aa	ac	c	c	t	tt	tt	gt	at	aa	tt	aa	gg	at	gt	tc	at	tc	1176
135	ttt	ggc	at	gt	at	ttt	gg	at	ttt	at	ttt	at	ttt	ttt	ttt	ttt	ttt	ttt	1236
137	at	ta	at	ca															1244
141	<210>	SEQ	ID	NO:	2														
143	<211>	LENGTH:	4700																
145	<212>	TYPE:	DNA																
147	<213>	ORGANISM:	Glycine max																
151	<220>	FEATURE:																	
153	<221>	NAME/KEY:	promoter																

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20:14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

155 <222> LOCATION: (1)..(1532)  
 159 <220> FEATURE:  
 161 <221> NAME/KEY: sig\_peptide  
 163 <222> LOCATION: (1533)..(1610)  
 167 <220> FEATURE:  
 169 <221> NAME/KEY: exon  
 171 <222> LOCATION: (1533)..(1751)  
 175 <220> FEATURE:  
 177 <221> NAME/KEY: exon  
 179 <222> LOCATION: (2383)..(2574)  
 183 <220> FEATURE:  
 185 <221> NAME/KEY: exon  
 187 <222> LOCATION: (3605)..(3769)  
 191 <220> FEATURE:  
 193 <221> NAME/KEY: exon  
 195 <222> LOCATION: (4033)..(4515)  
 199 <220> FEATURE:  
 201 <221> NAME/KEY: Intron  
 203 <222> LOCATION: (1752)..(2382)  
 207 <220> FEATURE:  
 209 <221> NAME/KEY: Intron  
 211 <222> LOCATION: (2575)..(3604)  
 215 <220> FEATURE:  
 217 <221> NAME/KEY: Intron  
 219 <222> LOCATION: (3770)..(4032)  
 222 <220> FEATURE:  
 224 <221> NAME/KEY: CDS  
 226 <222> LOCATION: (1533)..(1751)  
 230 <220> FEATURE:  
 232 <221> NAME/KEY: CDS  
 234 <222> LOCATION: (2383)..(2574)  
 238 <220> FEATURE:  
 240 <221> NAME/KEY: CDS  
 242 <222> LOCATION: (3605)..(3769)  
 246 <220> FEATURE:  
 248 <221> NAME/KEY: CDS  
 250 <222> LOCATION: (4033)..(4512)  
 254 <400> SEQUENCE: 2  
 255 tagataaaaa aatggat aattttctc agatgttatt tataactgtt ttttaatcag 60  
 257 aattaaaatt cctcttaat tatcgacata atttttttg gtgaatatta tcgacataat 120  
 259 tatttaatac aaattttat tgtacataga agtgcactt caatttaat attggagaac 180  
 261 agtacaaaaa cataaaaaaa ctgttattag aagaaaaaaa tatatggaaa aggttagcta 240  
 263 catatattag ctaaattatgt tttcttaatt ggcttatataa accctattgt actctttgta 300  
 265 atctcacctt tttcatttaa atacattct acttttaag ttcttatattt tctctcaatt 360  
 267 ttcttcgata aaccatgaaa tttaacatgg tatatcagcg ataccacca ctttggaaagc 420  
 269 catgtatggc tagtatggc agccaaaatt tgccctgggtt caagcaaagc aagtgtttat 480  
 271 atagatgtga ctttgttga ggaactcatg ccaatggtag tgattgtgaa actgagaaaa 540  
 273 ctaatttgg aatttgaat tatgtatcatt aaatactcct ctcctgacta cttcgtccc 600  
 275 tcaaatttgc accatcatta ttcccaaaa attgattac aatgcactaa ttaatgaatg 660

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20.14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

277	tttcttacat	tatcatat	tatcatatctga	cattttgttt	ttactttta	taataattat	720
279	tttaaaaagt	catacatgca	aataatttt	taatagttt	cagttaaatt	tttacagtaa	780
281	aatgcatga	aaattaaact	ttattttcc	aagtcatcat	ttagtcaaatt	cccaaaaacaa	840
283	tgattatttt	ttgcaaattga	atgtttattt	aacattttaa	tgttagctaa	ttaattctgg	900
285	ttatgggtgc	aatgttccaa	aacctaatgc	aagatcttag	caagtacata	catacatctaa	960
287	atttaaact	tatctttacg	caagagat	aaagattata	catctagtt	taaacattaa	1020
289	ctttgtttt	tgtgttaaaa	aacagtaaca	ttttcttaat	ttttagttagt	gacgtgctcc	1080
291	aaccatat	acgaagattt	taattggat	tcaagttcat	gaacttagta	aataagttt	1140
293	ggtttcagt	tttcaatttt	cattacaaca	tttatgtaaa	atatcaacgt	tttctgaaat	1200
295	ttgttgcttg	tgtgctccaa	ccacatttaa	gagattata	aaattaattt	tcaagaagat	1260
297	aatgattcct	actcttgctg	gccttaccat	agtacaataa	atccactcat	aatcaacaa	1320
299	gtcgtcgta	taggcaattt	ggcatcatat	cataaacaat	acgtacgtga	tattatctag	1380
301	tgtctctcag	tttactttat	gagaaattat	ttttctttaa	aaaaagttaa	ttaataaaaa	1440
303	catttgcgt	accgtgagtt	acaagaaatc	cgccgaattt	atctctataa	ataaaaggat	1500
305	ctatatgaga	gttaaaatca	tattaactca	aa atg ggt	tcc atg cgt	cta tta	1553
306				Met	Gly	Ser	
307				1	5		
309	gta	gtg	gca	ttg	ttg	tgt	1601
310	Val	Val	Ala	Leu	Leu	Cys	
311	10	15					
313	tct	tat	gct	cag	ctt	act	1649
314	Ser	Tyr	Ala	Gln	Leu	Thr	
315	25	30					
317	ctg	tcc	cct	att	gtg	ttt	
318	Leu	Phe	Pro	Ile	Val	Phe	
319	40	45					
321	ccc	cga	atc	ggg	gcc	agt	
322	Pro	Arg	Ile	Gly	Ala	Ser	
323	60	65					
325	gtt	caa	gtacgtactt	tttttttcc	ttccaaaatg	ccctgcata	1801
326	Val	Gln					
329	tgctttgttc	acctagaaaa	atgtgtttt	ttcaacgatc	ttacgtacgt	ttgtttgttt	1861
331	tgaaaaataa	atcagaaaga	gatcaagaaa	atagctagaa	agaaagcaac	gtttttttaa	1921
333	aaggtaattt	gtgtgagaaa	aatattttaa	ctgaagagaa	agaaattttaa	taagctttc	1981
335	ttgaatgata	tttacatgtc	tttataactt	aaagtccac	ttttcttta	agttgtgctt	2041
337	gaagaaaaaa	gatgttttc	agtttagttt	tgattatgc	taattatatt	ttaattttat	2101
339	taatttatac	tatatatcta	tttaccat	taattttac	tatatttcat	gatgacaaca	2161
341	gacaagtatt	ctaaagaggt	atcggtagat	gattaatttt	tttataaaaa	aatctttgc	2221
343	gtgtatagat	attctttat	aattggtgca	gaaactgtt	atgctaattt	caattatct	2281
345	tacattgatt	aactaatgc	tataatcaat	atttagttt	ggtataggag	acaaatcaag	2341
347	tgatctgaac	aaatthaagtt	gttatattt	cattgtgaca	g ggt	tgt	2397
348					75		
349							
351	gtt	ttg	ctg	aac	aac	act	2445
352	Val	Leu	Leu	Asn	Asn	Thr	
353	80	85					
355	cca	aat	atc	aac	tca	ata	2493
356	Pro	Asn	Ile	Asn	Ser	Ile	
357	95	100					
					105	110	

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20:14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

359 aca gcg gtg gaa aat agt tgt cca gac aca gtt tct tgt gct gat att	2541
360 Thr Ala Val Glu Asn Ser Cys Pro Asp Thr Val Ser Cys Ala Asp Ile	
115 120 125	
361 ctt gct att gca gct gaa ata gct tct gtt ctg gtaattaata actcctaatt	2594
363 ctt gct att gca gct gaa ata gct tct gtt ctg gtaattaata actcctaatt	
364 Leu Ala Ile Ala Ala Glu Ile Ala Ser Val Leu	
130 135	
365 aattcccaac cattaaaaag ttgcattgatt ggattcaaaa ttctatggta ttggggtttt	2654
367 gatataaatt tggatattaa ttgcactaaa aaaaattatc atatactttt aataaaaaaa	2714
369 atttatctaa tttaattttat tattaaaact atttttaaaa ttcaatccta actcttttt	2774
371 aatcgagca tggatcgcc caccacccgt atatcggtgg aagatgtat aaaaccattt	2834
373 aatcgagca tggatcgcc caccacccgt atatcggtgg aagatgtat aaaaccattt	2894
375 aattatggaa tggatcgcc caccacccgt atatcggtgg aagatgtat aaaaccattt	2954
377 atcatgttcg gcaagttac gttgtgtata attaatttga cttaatcaga taaaaaaca	3014
379 aatggacgca agccgggtgg tatagatatac actggcctgt agaatatgtg gttttcag	3074
381 tttaaataaa agctagctac tatattatatac actggcctgt agaatatgtg gttttcag	3134
383 cgtgatttat tgactgtgaa acatgtttcc acacacaggc ttagaaactc ctgcactata	3194
385 acatctccaa aatttgacta ttatattatg aagataattc atctatgtat ttcaactcta	3254
387 ttatataatgtatcatcgcc agtattaaaga attataatag tcaaataatag aagtatatcg	3314
389 gttaaatgtt gttgcattgtg cgacctgttt cgtgtaaaat gcttattcta tatactttt	3374
391 ttatattggaa aataacgtat aactaaaaac gaaagggtat catatagttt gactttatg	3434
393 ttagagagag acatcttaat ttggcatat gttaaataat taattacaat gcatacaca	3494
395 atatttatgc catatctaaa aaatgataaa atatcatagg tatactcaac tatatgatat	3554
397 cccataaca gaaattgtac ttcttcag gcaatgaact taacatttct gtttgctaaa	3610
399 aacaaacatc cactaaagt gttcaacat atttatgtaa taatttacag gga gga	
400 Gly Gly	3658
403 ggt cca gga tgg cca gtt cca tta gga aga agg gac agc tta aca gca	
404 Gly Pro Gly Trp Pro Val Pro Leu Gly Arg Arg Asp Ser Leu Thr Ala	
405 140 145 150 155	
407 aac cga acc ctt gca aat caa aac ctt cca gca cct ttc ttc aac ctc	3706
408 Asn Arg Thr Leu Ala Asn Gln Asn Leu Pro Ala Pro Phe Phe Asn Leu	
409 160 165 170	
411 act caa ctt aaa gct tcc ttt gct gtt caa ggt ctc aac acc ctt gat	3754
412 Thr Gln Leu Lys Ala Ser Phe Ala Val Gln Gly Leu Asn Thr Leu Asp	
413 175 180 185	
415 tta gtt aca ctc tca ggtatacata atcaattttt tatttgctat tagtagcaa	3809
416 Leu Val Thr Leu Ser	
417 190	
419 taaaaagtct ctgatacaga catatttataa taaatttttataat tctccataaaa catttataat	3869
421 aaaattatca atttatgtac taaaaatttataa tggatttggat ctctttcat ccaacttttta	3929
423 ctaaagttaa ggtgcataata atataaaaata aactatctt tttttttttt aaaaagattt	3989
425 aagataagtt aaagtctact tataatcat taatatatgt ata ggt ggt cat acg	4044
426 Gly Gly His Thr	
427 195	
429 ttt gga aga gct cgg tgc agt aca ttc ata aac cga tta tac aac ttc	4092
430 Phe Gly Arg Ala Arg Cys Ser Thr Phe Ile Asn Arg Leu Tyr Asn Phe	
431 200 205 210	
433 agc aac act gga aac cct gat cca act ctg aac aca aca tac tta gaa	4140
434 Ser Asn Thr Gly Asn Pro Asp Pro Thr Leu Asn Thr Thr Tyr Leu Glu	
435 215 220 225	
437 gta ttg cgt gca aga tgc ccc cag aat gca act ggg gat aac ctc acc	4188

## VERIFICATION SUMMARY

PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001

TIME: 10:20:15

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date

11/16/01